

Votre application d'IA générative est-elle digne de confiance ?

03 septembre 2026, 12:00

Mihail Mihaylov

Notre Trustworthiness Playbook vous aide à identifier les conditions nécessaires pour faire pleinement confiance à votre IA générative

L'intelligence artificielle générative (GenAI) évolue rapidement : d'outil expérimental, elle devient une technologie clé pour les applications industrielles et numériques. Les entreprises l'intègrent dans leurs produits, leurs flux de travail et leurs processus décisionnels pour gagner en efficacité, en flexibilité et en accessibilité. Mais son véritable potentiel dépend d'une question essentielle : « Pouvez-vous faire confiance à votre application ? » La réponse varie selon le secteur, l'application et l'entreprise. Notre Trustworthiness Playbook vous aide à identifier les critères essentiels à prendre en compte dans votre contexte.

Des exigences différentes selon le projet et le secteur

La confiance dans une application d'IA ne réside pas uniquement dans le modèle d'IA. L'ensemble du flux de travail, du flux de données et de la conception des processus est tout aussi important. Les exigences peuvent aussi varier fortement selon le contexte. Considérer chaque critère comme étant tout aussi critique risque donc de complexifier inutilement le développement.

Prenons un assistant GenAI interne qui aide un technicien de maintenance à trouver des informations dans de la documentation technique. Si une réponse erronée n'est pas souhaitable, le technicien garde cependant le contrôle et peut la vérifier.

Prenons maintenant un système d'IA qui établit des prévisions utilisées automatiquement pour soumettre des offres sur un marché de l'énergie. Les erreurs peuvent avoir des conséquences financières directes, tandis que des données d'entrée bruitées ou incomplètes peuvent affecter un processus automatisé. Dans les deux cas, il s'agit d'applications GenAI. Les critères nécessaires pour leur faire confiance, cependant, ne sont pas les mêmes.



Neuf dimensions pour une IA générative digne de confiance

Le playbook identifie neuf dimensions : **exactitude**, **conformité à la réglementation**, **explicabilité**, **robustesse**, **fiabilité**, **sécurité**, **transparence**, **équité** et **latence**.

L'exactitude vérifie si le système fournit la bonne réponse. **La fiabilité** évalue si son comportement reste cohérent dans des conditions normales. **La robustesse** examine s'il continue à fonctionner lorsque les données d'entrée sont bruitées, incomplètes ou différentes de celles pour lesquelles il a été conçu.

D'autres dimensions deviennent essentielles selon les circonstances. Un système accessible à des utilisateurs externes demandera davantage d'attention à la **sécurité**. Dans un environnement réglementé, la **conformité** et la **transparence** peuvent devenir prioritaires. **L'explicabilité** est importante lorsque les utilisateurs doivent comprendre et valider les recommandations. **L'équité** devient pertinente lorsque les résultats peuvent avoir des effets différents sur des individus ou des groupes. Et dans une application destinée aux opérateurs en temps réel, même une réponse exacte peut être inutile si la **latence** est trop élevée. Il n'est pas nécessaire de maximiser ces neuf dimensions. Concentrez-vous sur celles que votre application ne peut se permettre de négliger.

Un tableau décisionnel pour guider votre réflexion

C'est là que le **tableau décisionnel** du playbook peut jouer un rôle essentiel. Au lieu de partir de principes abstraits, il pose des questions pratiques sur votre application. Une réponse positive

permet d'identifier les dimensions qui doivent devenir incontournables.

Par exemple :

- L'IA intervient-elle dans des décisions opérationnelles, financières, médicales ou liées à la sécurité ? Concentrez-vous sur la **sécurité et l'explicabilité**
- Est-elle accessible à des utilisateurs externes ou exposée à des données d'entrée non fiables ? Donnez la priorité à la **sécurité**
- Les données d'entrée sont-elles bruitées, sujettes à des dérives ou incomplètes ? Maximisez la **robustesse**
- Des exigences réglementaires ou des exigences d'audit s'appliquent-elles ? Concentrez-vous sur la **conformité à la réglementation** et la **transparence**
- L'application doit-elle répondre en temps réel ? Investissez dans une **faible latence**
- Les utilisateurs ou les auditeurs doivent-ils comprendre son comportement et ses limites ? Donnez la priorité à la **transparence**
- L'application réalise-t-elle des prédictions, des prévisions, de l'automatisation ou du contrôle ? **L'exactitude et la fiabilité** sont alors vos principales préoccupations

Ce tableau décisionnel vous aide à transformer une vaste discussion sur la gestion des risques liés à l'IA en un instrument pratique que les équipes d'ingénierie, produit et innovation peuvent utiliser directement dans leurs projets.

Exemple : assistance interactive aux opérateurs

Le playbook présente également des exemples concrets pour illustrer cette approche. Prenons un environnement de production où les collaborateurs utilisent un assistant basé sur l'IA pour réaliser des tâches plus complexes de manière autonome. Le système peut combiner différentes sources d'information, comme des documents techniques, des images et des questions orales. Il peut aussi interagir avec des outils externes afin de fournir aux opérateurs des informations et recommandations adaptées au contexte.

Pour une application d'assistance aux opérateurs, **l'exactitude, la fiabilité et une faible latence** constituent des points de départ logiques. Les collaborateurs ont besoin d'une assistance correcte et cohérente et s'attendent à recevoir des réponses en quelques secondes. Le tableau décisionnel fait également apparaître d'autres priorités.

L'assistant génère des recommandations et du contenu, ce qui fait de **l'explicabilité** un aspect important. L'interaction avec des opérateurs moins à l'aise avec la technologie exige également une attention particulière à la **sécurité**. Et comme les données vocales peuvent être bruitées et les transcriptions imparfaites, la **robustesse** devient essentielle. Le système doit en effet continuer à fournir une assistance utile même lorsque ses données d'entrée ne sont pas optimales.

L'évaluation fait ainsi ressortir six exigences incontournables : **exactitude, fiabilité, explicabilité, sécurité, robustesse et faible latence**. **La transparence et la conformité à la réglementation** restent utiles, mais ne constituent pas nécessairement des priorités pour cette application précise.

Ces exigences peuvent ensuite être traduites en choix de conception pratiques. La fiabilité peut, par exemple, être renforcée en structurant les interactions avec les outils et les mécanismes de validation. La sécurité peut être améliorée en limitant le système à des actions prédéfinies et en maintenant les opérateurs dans la boucle de décision.

Les conditions pour une IA générative digne de confiance

Même une application GenAI très performante peut s'avérer inutile si les conditions nécessaires à la confiance ne sont pas réunies. Les efforts pour rendre l'IA générative digne de confiance resteront insuffisants si les données d'entrée sont de mauvaise qualité, si les résultats ne sont pas validés, si les utilisateurs comprennent mal les limites du système ou si le processus accorde trop d'autonomie à l'IA.

À l'inverse, de nombreuses mesures permettent de renforcer la confiance. Notre playbook en répertorie un grand nombre, avec des recommandations pratiques plutôt que de simples principes théoriques. Ancrer les résultats dans des données de référence, ajouter des contrôles de validation, standardiser les prompts, surveiller les dérives, définir une logique de repli, limiter les fonctionnalités sensibles ou maintenir une intervention humaine : autant de mesures concrètes et directement applicables que vous trouverez dans ce document.

En résumé

Une IA générative digne de confiance exige plus qu'un bon modèle d'IA. La notion de confiance dépend de la tâche, des données, des utilisateurs, du degré d'autonomie, de l'environnement opérationnel, des conséquences d'une erreur et des réglementations applicables.

Le « Trustworthiness Playbook » de Sirris et Flanders Make aide les équipes à réaliser cette évaluation de manière systématique. Il combine neuf dimensions d'une IA digne de confiance avec des catégories de cas d'usage reconnaissables, un tableau décisionnel pratique et des stratégies concrètes d'amélioration.

Rendez la fiabilité concrète pour votre application GenAI

Quels aspects de la fiabilité sont vraiment importants pour votre cas d'usage GenAI ? Téléchargez le Trustworthiness Playbook pour identifier vos priorités et découvrir des mesures concrètes pour y répondre.

[Téléchargez le playbook](#)

Des questions sur la fiabilité de la GenAI ?

Vous avez encore des questions sur la fiabilité ou sur la manière d'appliquer ces principes à votre cas d'usage GenAI ? Mihail Mihaylov, Senior Engineer AI & GenAI chez Sirris, se fera un plaisir d'en discuter avec vous.

[Contactez notre expert](#)

Auteurs

Profile picture for user mihail.mihaylov@sirris.

Mihail Mihaylov